AD-A077 654 WESTERN MICHIGAN UNIV KALAMAZOO DEPT F MATHEMATICS F/G 12/1
A WEIGHTED DISPERSION FUNCTION FOR ES MATION IN LINEAR MODELS,(U)
OCT 79 G L SIEVERS N00014-78-C-0637
UNCLASSIFIED MATHEMATICS-62 NL

1 OF 1

AD
A077654

END
DATE
FILMED
12-79
DDC

DEPARTMENT OF MATHEMATICS

WESTERN MICHIGAN UNIVERSITY
Kalamazoo, Michigan
49008

~~WESTERN MICHIGAN UNIVERSITY~~

14 Mathematics ~~Report~~ 62

~~October, 1979~~

6

A WEIGHTED DISPERSION FUNCTION FOR ESTIMATION IN LINEAR MODELS

by

10 GERALD L. SIEVERS

11 Oct 79

12 39

15 N00014-78-C-0637

407959

SUMMARY

Robust estimates for the parameters in the general linear model are proposed which are based on weighted rank statistics. The method is based on the minimization of a dispersion function defined by a weighted Gini's mean difference. An asymptotic distribution of the estimate is derived. Some examples are discussed which point out that the ranking can be based on a restricted set of comparisons and still retain high efficiency.

Key words: Robust estimates, linear models, weighted rank statistics, dispersion function, Gini mean difference

1. INTRODUCTION

Consider the linear model

$$\underset{\sim}{Y} = \beta_o \underset{\sim}{1} + \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{e} , \tag{1.1}$$

where $\underset{\sim}{Y} = (Y_1, \ldots, Y_n)'$ is an $n \times 1$ random vector, $\underset{\sim}{1}$ is an $n \times 1$ vector with each element equal to one, $\underset{\sim}{X} = (x_{ij})$ is an $n \times p$ design matrix, $\underset{\sim}{\beta} = (\beta_1, \ldots, \beta_p)'$ is a $p \times 1$ vector of parameters and $\underset{\sim}{e} = (e_1, \ldots, e_n)'$ is an $n \times 1$ vector of random errors. Assume that $e_1, e_2, \ldots, e_n$ are independent with a common distribution having density function f. The residuals are given by $\underset{\sim}{Z} = (Z_1, \ldots, Z_n)'$ where

$$\underset{\sim}{Z} = \underset{\sim}{Z}(\underset{\sim}{\beta}) = \underset{\sim}{Y} - \underset{\sim}{X}\underset{\sim}{\beta} .$$

Methods of estimation of $\underset{\sim}{\beta}$ are typically based on some principle of making the residuals small. The classical least-squares approach is to minimize the sum of squares of the residuals. The resulting estimate is optimal under normality assumptions. However, the least-squares estimate is not robust in the face of departures from the model. It can be inefficient when the error terms follow a non-normal distribution and it can be very sensitive to outliers and high leverage points in the design matrix. These problems with least-squares estimates have spurred the development of other types of estimates which are more robust.

There has been considerable work in recent years on the M-estimate approach and on the method of minimizing the sum of the absolute values of the residuals (Least Absolute Deviation estimates). Methods based on rank statistics have also been proposed. With regard to the rank statistic approach, basic material can be found in the papers of Jurečková (1969), (1971); Koul (1970), (1971); Adichie (1978) and Sen-Puri (1977).

Jaeckel (1972) has discussed the value in using a dispersion function and in defining the estimates to be the values of the parameters that minimize the dispersion of the residuals. He showed how estimates based on linear rank statistics can arise with a suitable choice of dispersion function. This approach has been further extended by Hettmansperger and McKean (1976), (1977), (1978b).

This paper will examine the estimate of $\tilde{\beta}$ that arises with a dispersion function defined as a weighted Gini's mean difference. Gini's mean difference is a familiar measure of dispersion and it has been proposed for the linear model problem by Wainer and Thissen (1976). The use of weights adds greater flexibility. The asymptotic theory of the partial derivatives of the dispersion function will be examined and an asymptotic linearity result is given. This dispersion function is shown to be asymptotically, locally quadratic. These results are used to establish the asymptotic distribution of the proposed estimate of $\tilde{\beta}$. The paper concludes with some comments on the weights and some applications. Proofs of the theorems can be found in the Appendix.

2. THE DISPERSION FUNCTION

Consider the dispersion function

$$D = D(\underset{\sim}{\beta}) = \sum_{i<j} b_{ij} |Z_i - Z_j|, \tag{2.1}$$

where the $b_{ij} \geq 0$, $1 \leq i < j \leq n$ are a given set of weights. Each pair of residuals is compared by the absolute difference and the weight that is attached can reflect the importance of the comparison. Note that the weights can depend on the design matrix $\underset{\sim}{X}$. It is possible to have some of the weights equal to zero and this will drop some pairwise comparisons from consideration. The special case of equal weights, $b_{ij} \equiv 1$, gives rise to Gini's mean difference. Hettmansperger and McKean (1978a) have shown that this dispersion function is equivalent to Jaeckel's dispersion function with Wilcoxon scores.

The dispersion function D can be expressed in another form. Let $(R_1, \ldots, R_n)$ denote the ranks of the residuals; that is, R_i is the rank of Z_i in the set $\{Z_1, \ldots, Z_n\}$, $1 \leq i \leq n$. Let $\text{sgn}(v) = +1, 0, -1$ as v is >0, $=0$, <0. Extend the definition of the weights b_{ij} to all subscripts $i, j = 1, \ldots, n$ by using $b_{ji} = b_{ij}$ and $b_{ii} = 0$. Then, using $|v| = v\,\text{sgn}(v)$, some manipulation shows

$$D = \sum_{i=1}^{n} B_i Z_i, \tag{2.2}$$

with $B_i = \sum_{j \neq i} b_{ij} \,\text{sgn}(Z_i - Z_j)$, $i = 1, \ldots, n$.

The coefficients B_i are random with B_i depending on the rank of Z_i and also on the subscripts of the residuals that are less than Z_i. In the special case $b_{ij} \equiv 1$, $B_i = 2R_i - (n + 1)$.

Another dispersion function which is similar to D is

$$D^* = \sum_{i<j} b_{ij} |Z_{(i)} - Z_{(j)}| ,$$

in which the weights correspond to the ordered residuals $Z_{(1)} \leq \cdots \leq Z_{(n)}$. It can be shown that

$$D^* = \sum_{i=1}^{n} B^*_{R_i} Z_i = \sum_{i=1}^{n} B^*_i Z_{(i)} ,$$

$$\text{where} \quad B^*_i = \sum_{j=1}^{i-1} b_{ij} - \sum_{j=i+1}^{n} b_{ij} .$$

In this form, it can be seen that D^* is equivalent to Jaeckel's dispersion function with the B^*_i serving as the score function. If $b_{ij} \equiv 1$, then $D = D^*$. They are not equal in general. This shows that the weights used in D serve a different purpose than the score function used by Jaeckel.

3. PARTIAL DERIVATIVES OF D

To estimate the parameter $\underset{\sim}{\beta}$, consider using a point in the parameter space which minimizes the dispersion function $D(\beta)$ of (2.1). This function is nonnegative, piecewise linear and convex. Various numerical methods, including linear programming algorithms, can be used to determine an estimate. The solution is not unique in general. However, under some conditions, it follows from the work in section 5 that the diameter of the set of solutions tends to zero asymptotically.

The partial derivatives of D should be (approximately) equal to zero at the minimum. Using form (2.2), these derivatives are

$$\partial D/\partial\beta_k = -\sum_{i=1}^{n} B_i x_{ik} \tag{3.1}$$

for $k = 1, \ldots, p$, at points $\underset{\sim}{\beta}$ where they exist. Another form of the derivatives, that can be seen by writing $D = \sum_{i<j} b_{ij} \, \mathrm{sgn}(Z_j - Z_i)(Z_j - Z_i)$ is

$$\begin{aligned} \partial D/\partial\beta_k &= -\sum_{i<j} b_{ij} \, \mathrm{sgn}(Z_j - Z_i)(x_{jk} - x_{ik}) \\ &= -2\sum_{i<j} b_{ij}(x_{jk} - x_{ik})\phi(Z_i, Z_j) + \sum_{i<j} b_{ij}(x_{jk} - x_{ik}) , \end{aligned} \tag{3.2}$$

where $\phi(u,v) = (\mathrm{sgn}(v - u) + 1)/2 = 0, 1/2, 1$ as $u > v$, $u = v$, $u < v$. In this form, the derivatives can be seen to depend on the rank order of the residuals. They involve a general type of random variable of the form

$\sum_{i<j} a_{ij}\, \phi(Z_i, Z_j)$, which will be considered in more detail in the next section.

The form of the derivatives in (3.2) can be changed to

$$\partial D/\partial\beta_k = \sum_{i<j} b_{ij} |x_{jk} - x_{ik}| \, \mathrm{sgn}(x_{jk} - x_{ik}) \mathrm{sgn}(Z_j - Z_i),$$

for $k = 1, \ldots, p$. This is a "weighted" Kendall's tau random variable for Z vs x_k. Thus when the partial derivatives are zero, the residuals are uncorrelated with the independent variables in this nonparametric sense. This is directly analogous to the least-squares approach where the least-squares estimate of $\tilde{\beta}$ can be defined by specifying that the residuals be uncorrelated (Pearson product moment correlation) with the independent variables.

4. A GENERAL CLASS OF RANDOM VARIABLES

In this section a general class of random variables, related to the derivatives of the dispersion function, is defined. An asymptotic normality result is given with the proof delayed until the Appendix.

For each $k = 1, \ldots, p$, let a set of constants $\{a_{ij}(k): 1 \leq i < j \leq n\}$ be given. Let

$$a_{i\cdot}(k) = \sum_{j=i+1}^{n} a_{ij}(k) \qquad \text{for } i = 1, \ldots, n-1$$

$$a_{\cdot j}(k) = \sum_{i=1}^{j-1} a_{ij}(k) \qquad \text{for } j = 2, \ldots, n$$

$$a_{n\cdot}(k) = 0, \quad a_{\cdot 1}(k) = 0, \quad a_{\cdot\cdot}(k) = \sum_{i<j} a_{ij}(k)$$

$$A_i(k) = a_{\cdot i}(k) - a_{i\cdot}(k) .$$

For asymptotic purposes, a sequence of these constants is needed, indexed on $n = 1, 2, \ldots,$ but this dependence on n will be suppressed in the notation. In a similar fashion, the dependence on n of other quantities will not be indicated in the notation.

ASSUMPTION (A_1): For each $k = 1, \ldots, p$

$$\frac{\sum_{i=1}^{n} A_i^2(k)}{\max_{1 \leq i \leq n} A_i^2(k)} \to \infty \quad \text{as } n \to \infty .$$

ASSUMPTION (A_2): For each $k = 1, \ldots, p$

$$\frac{\sum_{i<j} a_{ij}^2(k)}{\sum_{i=1}^{n} A_i^2(k)} \to 0 \qquad n \to \infty .$$

Define the random variables

$$T_k = T_k(\underset{\sim}{\beta}) = \sum_{i<j} a_{ij}(k)\phi(Z_i, Z_j) ,$$

for $k = 1, \ldots, p$, where $\phi(u, v) = 0, 1/2, 1$ as $u > v$, $u = v$, $u < v$. Let $\underset{\sim}{T} = \underset{\sim}{T}(\underset{\sim}{\beta}) = (T_1, \ldots, T_p)'$ be the $p \times 1$ vector of these random variables. Note that this type of random variable arises in the derivatives of the dispersion function in (3.2) with the correspondence $a_{ij}(k) = b_{ij}(x_{jk} - x_{ik})$. Specifically,

$$\partial D/\partial \beta_k = -2T_k(\underset{\sim}{\beta}) + a_{..}(k) \quad . \tag{4.1}$$

In order to consider the asymptotic distribution of $\underset{\sim}{T}(\underset{\sim}{\beta})$, the following notation is introduced.

Specify a sequence of parameter values, contiguous to $\underset{\sim}{0}$ by assuming $\underset{\sim}{\beta} = \underset{\sim}{\Delta}/\sqrt{n}$, where $\underset{\sim}{\Delta}' = (\Delta_1, \ldots, \Delta_p)$ is a fixed vector.

Let the centered design matrix be $\underset{\sim}{X}_c = (\underset{\sim}{I}_n - (1/n)\underset{\sim}{J}_n)\underset{\sim}{X}$, where $\underset{\sim}{I}_n$ is an $n \times n$ identity matrix and $\underset{\sim}{J}_n$ is an $n \times n$ matrix of "ones". Let $\bar{x}_k$ be the average of the k^{th} column of $\underset{\sim}{X}$, $k = 1, \ldots, p$.

Let $\underset{\sim}{A}_n$ be an $n \times p$ matrix with $(i, k)^{th}$ element equal to $A_i(k)$, $i = 1, \ldots, n$, $k = 1, \ldots, p$, and let $\underset{\sim}{V}_n = \underset{\sim}{A}_n' \underset{\sim}{A}_n$. Let $\underset{\sim}{a}_{..}' = (a_{..}(1), \ldots, a_{..}(p))$ and let $\underset{\sim}{\mu}_n' = (\mu_n(1), \ldots, \mu_n(p)) = (\int f^2)\, \underset{\sim}{A}_n' \underset{\sim}{X}_c \underset{\sim}{\beta} + (1/2)\underset{\sim}{a}_{..}$.

<u>ASSUMPTION (A_3)</u>: For $k = 1, \ldots, p$

$$\max_{1 \le i \le n} \frac{|x_{ik} - \bar{x}_k|}{\sqrt{n}} \to 0 \quad \text{as} \quad n \to \infty \quad .$$

<u>ASSUMPTION (A_4)</u>:

$$(1/n)\underset{\sim}{X}_c' \underset{\sim}{X}_c \to \underset{\sim}{\Sigma} \qquad \text{as} \quad \to \infty \quad ,$$

where $\underset{\sim}{\Sigma}$ is a $p \times p$, positive definite matrix.

<u>ASSUMPTION (A_5)</u>: There exists a sequence of constants $\{\gamma_n\}$ such that

$$\gamma_n \underset{\sim}{V}_n \to \underset{\sim}{V} \quad \text{as} \quad n \to \infty \quad ,$$

where $\underset{\sim}{V}$ is a $p \times p$, positive definite matrix.

<u>**THEOREM 4.1**</u> Assume the error density f is absolutely continuous and $\int (f'/f)^2 f \, dx < \infty$. Let assumptions (A_1), (A_2), (A_3), (A_4) and (A_5) hold. Then, if $\underset{\sim}{\beta} = \underset{\sim}{\Delta}/\sqrt{n}$,

$$\gamma_n^{1/2}(\underset{\sim}{T}(\underset{\sim}{0}) - \underset{\sim}{\mu}_n) \xrightarrow{\mathcal{L}} N(\underset{\sim}{0}, (1/12)\underset{\sim}{V})$$

as $n \to \infty$

The notation " $\xrightarrow{\mathcal{L}}$ " reads "converges in distribution". A translation property of the result can be noted since $\underset{\sim}{T}(\underset{\sim}{\beta}_1)|_{\underset{\sim}{\beta}_2} \overset{\mathcal{L}}{=} \underset{\sim}{T}(\underset{\sim}{0})|_{\underset{\sim}{\beta}_2 - \underset{\sim}{\beta}_1} \overset{\mathcal{L}}{=} \underset{\sim}{T}(\underset{\sim}{\beta}_1 - \underset{\sim}{\beta}_2)|_{\underset{\sim}{0}}$. Here $\underset{\sim}{T}(\underset{\sim}{\beta}_1)|_{\underset{\sim}{\beta}_2}$ refers to the distribution of $T(\underset{\sim}{\beta}_1)$ when $\underset{\sim}{\beta} = \underset{\sim}{\beta}_2$.

5. ASYMPTOTIC LINEARITY

In this section, a local, asymptotic linearity result is given for $\underset{\sim}{T}(\underset{\sim}{\beta})$. The proof can be found in the Appendix.

Let $\underset{\sim}{\Delta} = (\Delta_1, \ldots, \Delta_p)'$ denote a $p \times 1$ vector and let $c > 0$ be given. Define a set $\mathcal{D} = \{\underset{\sim}{\Delta}: -c \leq \Delta_k \leq c, \; k = 1, \ldots, p\}$.

Let a $p \times p$ matrix $\underset{\sim}{C}_n$ be defined with $(k, \ell)^{th}$ element $c_{k\ell} = -(\int f^2) \sum_{i<j} a_{ij}(k)(x_{j\ell} - x_{i\ell})$. Let

$$(5.1) \qquad \underset{\sim}{R}(\underset{\sim}{\Delta}) = n^{-3/2}[\underset{\sim}{T}(\underset{\sim}{\Delta}/\sqrt{n}) - \underset{\sim}{T}(\underset{\sim}{0}) - \underset{\sim}{C}_n(\underset{\sim}{\Delta}/\sqrt{n})] \; .$$

Let $G(y) = P(e_1 - e_2 \leq y)$ denote the cdf of the difference of independent random variables, each with density f.

Let $\|\cdot\|$ denote Euclidean distance.

ASSUMPTION (A_6): The cdf G has a density $g = G'$ and $g(y)$ is continuous at $y = 0$.

ASSUMPTION (A_7): For each $k = 1, \ldots, p$

$$\frac{\sum_{i<j} a_{ij}^2(k)}{\binom{n}{2}} \quad \text{is bounded as} \quad n \to \infty \; .$$

LEMMA 5.1 Let assumptions (A_3), (A_4), (A_6) and (A_7) hold. If $\underset{\sim}{\beta} = \underset{\sim}{0}$, then $\underset{\sim}{R}(\underset{\sim}{\Delta}) \xrightarrow{P} 0$, uniformly in $\underset{\sim}{\Delta} \in \mathcal{D}$. (That is, for all $\varepsilon > 0$ and $\delta > 0$, there exists N such that $P(\|\underset{\sim}{R}(\underset{\sim}{\Delta})\| \geq \varepsilon) \leq \delta$ for all $n \geq N$ and all $\underset{\sim}{\Delta} \in \mathcal{D}$.)

The lemma shows that $\tilde{T}(\tilde{\beta})$ can be approximated by the linear function $\tilde{T}(\tilde{0}) + \tilde{C}_n\tilde{\beta}$ asymptotically for $\tilde{\beta}$ sufficiently near zero, however the result is not strong enough for the application needed. A stronger result is given in the following theorem.

<u>THEOREM 5.1</u> Let assumptions (A_3), (A_4), (A_6) and (A_7) hold. If $\tilde{\beta} = \tilde{0}$, then $\sup_{\tilde{\Delta}\in\mathcal{D}} \|\tilde{R}(\tilde{\Delta})\| \xrightarrow{P} 0$ as $n \to \infty$.

6. DISTRIBUTION OF $\hat{\beta}$

The estimate of β has been defined to be a point in the parameter space, say $\hat{\beta}_n$, which minimizes the dispersion function $D(\beta)$ of (2.1). The set of solutions to this minimization problem is bounded as seen by the following lemma. The proof of this lemma is exactly the same as that of Theorem 2 of Jaeckel (1972).

LEMMA 6.1 If the centered design matrix X_c is of full rank p, then $\{\beta: D(\beta) \leq D_o\}$ is bounded for any number D_o.

In order to deal with the asymptotic distribution of $\hat{\beta}_n$, it is convenient to work with $\hat{\Delta}_n = \sqrt{n}\,\hat{\beta}_n$ and define

$$D^*(\Delta) = (1/n)D(\Delta/\sqrt{n}) .$$

Then $\hat{\Delta}_n$ minimizes $D^*(\Delta)$.

To match the dispersion function D^* to the T vector of section 4, use the correspondence

$$a_{ij}(k) = b_{ij}(x_{jk} - x_{ik}) . \tag{6.1}$$

Then from formula (4.1), the vector of partial derivatives is

$$\frac{\partial D^*(\Delta)}{\partial \Delta} = n^{-3/2}[-2T(\Delta/\sqrt{n}) + a_{..}] .$$

With (6.1) we can give more definite expressions for some matrices that were defined earlier. The elements of the matrix $\underset{\sim}{A}_n$ of section 4 are linear functions of the elements x_{ij} of the design matrix and with some manipulation, write

$$\underset{\sim}{A}_n = \underset{\sim}{B}_n \underset{\sim}{X} ,$$

where $\underset{\sim}{B}_n$ is an $n \times n$ symmetric matrix involving the weights of the dispersion function. Specifically, define the $(i, j)^{th}$ element of $\underset{\sim}{B}_n$ to be $-b_{ij}$ if $i < j$ and $-b_{ji}$ if $i > j$. The i^{th} diagonal element of $\underset{\sim}{B}_n$ is $b_i = \sum_{j \neq i} b_{ij}$. Thus $\underset{\sim}{B}_n$ has the negatives of the dispersion function weights for its off-diagonal elements and positive diagonal elements determined so that the row sums are zero. Also write

$$\underset{\sim}{V}_n = \underset{\sim}{A}'_n \underset{\sim}{A}_n = \underset{\sim}{X}' \underset{\sim}{B}_n \underset{\sim}{B}_n \underset{\sim}{X} .$$

Again with (6.1), the matrix $\underset{\sim}{C}_n$ of section 5 will have $(k, \ell)^{th}$ element $c_{k\ell} = -(\int f^2) \sum_{i<j} b_{ij}(x_{jk} - x_{ik})(x_{j\ell} - x_{i\ell})$. With some further manipulation, it follows that

$$\underset{\sim}{C}_n = -(\int f^2)\, \underset{\sim}{X}' \underset{\sim}{B}_n \underset{\sim}{X} .$$

The matrices $\underset{\sim}{V}_n$ and $\underset{\sim}{C}_n$ have been expressed directly in terms of the design matrix $\underset{\sim}{X}$ and the weight matrix $\underset{\sim}{B}_n$. Since the row (and column) sums of $\underset{\sim}{B}_n$ are zero, the centered design matrix could be used with

$$\underset{\sim}{V}_n = \underset{\sim}{X}'_c \underset{\sim}{B}_n \underset{\sim}{B}_n \underset{\sim}{X}_c$$

$$\underset{\sim}{C}_n = -(\int f^2)\underset{\sim}{X}'_c \underset{\sim}{B}_n \underset{\sim}{X}_c \quad .$$

<u>ASSUMPTION (A_8)</u>: With (6.1) holding,

$$(1/n^2)\underset{\sim}{C}_n \to \underset{\sim}{C} \quad \text{as } n \to \infty \quad ,$$

where $\underset{\sim}{C}$ is a $p \times p$ matrix of full rank.

Now define a quadratic function of $\underset{\sim}{\Delta}$ to use as an approximation to $D^*(\underset{\sim}{\Delta})$ by

$$Q(\underset{\sim}{\Delta}) = -\underset{\sim}{\Delta}' \underset{\sim}{C} \underset{\sim}{\Delta} + n^{-3/2}[\underset{\sim}{a}_{..} - 2\underset{\sim}{T}(0)]' \underset{\sim}{\Delta} + D(\underset{\sim}{0}) \quad .$$

Then

$$\frac{\partial Q(\underset{\sim}{\Delta})}{\partial \underset{\sim}{\Delta}} = -2\underset{\sim}{C} \underset{\sim}{\Delta} + n^{-3/2}[\underset{\sim}{a}_{..} - 2\underset{\sim}{T}(\underset{\sim}{0})] \quad .$$

<u>LEMMA 6.2</u> Let assumption (A_8) hold along with the assumptions of Theorem 5.1. Then

$$\sup_{\underset{\sim}{\Delta} \in \mathcal{D}} n^{-3/2} \| \underset{\sim}{T}(\underset{\sim}{\Delta}/\sqrt{n}) - \underset{\sim}{T}(\underset{\sim}{0}) - n^{3/2}\underset{\sim}{C}\underset{\sim}{\Delta} \| \xrightarrow{P} 0 \quad \text{as } n \to \infty \quad .$$

<u>LEMMA 6.3</u> Let the assumptions of Lemma 6.2 hold. Then

$$\sup_{\underset{\sim}{\Delta} \in \mathcal{D}} \left\| \frac{\partial Q(\underset{\sim}{\Delta})}{\partial \underset{\sim}{\Delta}} - \frac{\partial D^*(\underset{\sim}{\Delta})}{\partial \underset{\sim}{\Delta}} \right\| \xrightarrow{P} 0 \quad \text{as } n \to \infty \quad .$$

<u>LEMMA 6.4</u> Let the assumptions of Lemma 6.2 hold. Then

$$\sup_{\underset{\sim}{\Delta} \in \mathcal{D}} |Q(\underset{\sim}{\Delta}) - D^*(\underset{\sim}{\Delta})| \xrightarrow{P} 0 \quad \text{as } n \to \infty$$

Having established that $Q(\underset{\sim}{\Delta})$ can be used to approximate $D^*(\underset{\sim}{\Delta})$, it will next be shown that $\underset{\sim}{\Delta}$ values minimizing these functions must be close enough to have the same asymptotic distribution.

On setting the partial derivatives equal to zero, $Q(\underset{\sim}{\Delta})$ is minimized at the point

$$\underset{\sim}{\Delta}^*_n = n^{-3/2} \underset{\sim}{C}^{-1}[(1/2)\underset{\sim}{a}_{..} - \underset{\sim}{T}(\underset{\sim}{0})] .$$

<u>THEOREM 6.1</u> Let assumption (A_8) hold along with the assumptions of Theorem 5.1. Let assumption (A_5) hold with $\gamma_n = n^{-3}$. Then $\hat{\underset{\sim}{\Delta}}_n$ and $\underset{\sim}{\Delta}^*_n$ are asymptotically equivalent. (See definition on page 1453 of Jaeckel (1972).)

Now if Theorems 4.1 and 6.1 hold, when $\underset{\sim}{\beta} = \underset{\sim}{0}$

$$n^{-3/2}[\underset{\sim}{T}(\underset{\sim}{0}) - (1/2)\underset{\sim}{a}_{..}] \xrightarrow{\mathcal{L}} N(\underset{\sim}{0}, (1/12)\underset{\sim}{V}) \quad \text{as} \quad n \to \infty$$

It follows that $\underset{\sim}{\Delta}^*_n$ is asymptotically $N(\underset{\sim}{0}, (1/12)\underset{\sim}{C}^{-1} \underset{\sim}{V} \underset{\sim}{C}^{-1})$ and the following result is immediate.

<u>COROLLARY 6.1</u> Under the assumptions of Theorems 4.1 and 6.1, when $\underset{\sim}{\beta} = \underset{\sim}{0}$

$$\sqrt{n}\ \hat{\underset{\sim}{\beta}}_n = \hat{\underset{\sim}{\Delta}}_n \quad \text{is asymptotically} \quad N(\underset{\sim}{0}, (1/12)\underset{\sim}{C}^{-1} \underset{\sim}{V} \underset{\sim}{C}^{-1}) .$$

Note that from the translation invariance of the estimate,

$$\sqrt{n}(\hat{\underset{\sim}{\beta}}_n - \underset{\sim}{\beta})\big|_{\underset{\sim}{\beta}} \overset{\mathcal{L}}{=} \sqrt{n}\ \hat{\underset{\sim}{\beta}}_n\big|_{\underset{\sim}{0}} .$$

If $n\,\underset{\sim}{C}_n^{-1}\,\underset{\sim}{V}_n\,\underset{\sim}{C}_n^{-1}$ is used as an approximation to $\underset{\sim}{C}^{-1}\,\underset{\sim}{V}\,\underset{\sim}{C}^{-1}$, the asymptotic covariance matrix of $\hat{\underset{\sim}{\beta}}_n$ is approximately

$$
\begin{aligned}
\operatorname{cov}(\hat{\underset{\sim}{\beta}}_n) &= (1/12)\,\underset{\sim}{C}_n^{-1}\,\underset{\sim}{V}_n\,\underset{\sim}{C}_n^{-1} \\
&= (1/12(\textstyle\int f^2)^2)(\underset{\sim}{X}'\,\underset{\sim}{B}_n\,\underset{\sim}{X})^{-1}(\underset{\sim}{X}'\,\underset{\sim}{B}_n\,\underset{\sim}{B}_n\,\underset{\sim}{X})(\underset{\sim}{X}'\,\underset{\sim}{B}_n\,\underset{\sim}{X})^{-1}\,.
\end{aligned}
\tag{6.2}
$$

Note that $\underset{\sim}{X}_c$ can be used in place of $\underset{\sim}{X}$ in this formula. It should be emphasized that $\operatorname{cov}(\hat{\underset{\sim}{\beta}}_n)$ is not the exact covariance matrix. In spite of this, formula (6.2) may prove useful in examing the effect of different weight matrices on the estimate.

For the special case $b_{ij} \equiv 1$, the unweighted case, $\underset{\sim}{B}_n = n(\underset{\sim}{I}_n - (1/n)\underset{\sim}{J}_n)$ and

$$
\operatorname{cov}(\hat{\underset{\sim}{\beta}}_n) = (1/12(\textstyle\int f^2)^2)(\underset{\sim}{X}_c'\,\underset{\sim}{X}_c)^{-1}\,.
\tag{6.3}
$$

Note that $\operatorname{cov}(\hat{\underset{\sim}{\beta}}_n)$ depends on $\underset{\sim}{B}_n$ through the matrix $\underset{\sim}{H} = \underset{\sim}{B}_n\,\underset{\sim}{X}$ since $[(\underset{\sim}{X}'\,\underset{\sim}{B}_n\,\underset{\sim}{X})(\underset{\sim}{X}'\,\underset{\sim}{B}_n\,\underset{\sim}{B}_n\,\underset{\sim}{X})^{-1}(\underset{\sim}{X}'\,\underset{\sim}{B}_n\,\underset{\sim}{X})]^{-1} = [\underset{\sim}{X}'\,\underset{\sim}{H}(\underset{\sim}{H}'\,\underset{\sim}{H})^{-1}\,\underset{\sim}{H}'\,\underset{\sim}{X}]^{-1}$. The matrix $\underset{\sim}{H}(\underset{\sim}{H}'\,\underset{\sim}{H})^{-1}\,\underset{\sim}{H}'$ is a projection matrix, the projection being into the column space of $\underset{\sim}{H}$.

7. THE WEIGHT MATRIX B_n

In the unweighted case, $b_{ij} \equiv 1$ and $B_n = n(I_n = (1/n)J_n)$. This is a benchmark case because of its simplicity. It also yields the highest asymptotic efficiency as can be seen by expressing the difference of the covariance matrices (6.2) and (6.3) as (excepting the constant multiple)

$$(X_c' B_n X_c)^{-1}(X_c' B_n B_n X_c)(X_c' B_n X_c)^{-1} - (X_c' X_c)^{-1} = M M'$$

where $M = (X_c' B_n X_c)^{-1} X_c' B_n - (X_c' X_c)^{-1} X_c'$. The matrix $M M'$ is positive semidefinite and as a result the trace and determinant of (6.2) cannot be less than that of (6.3).

Formulas (6.2) and (6.3) are equal if and only if $M = 0$ or $(X_c' B_n X_c)^{-1} X_c' B_n = (X_c' X_c)^{-1} X_c'$. Equivalently, $B_n X_c = X_c (X_c' X_c)^{-1} X_c' B_n X_c$. Since $X_c (X_c' X_c)^{-1} X_c'$ is the projection map into the column space of X_c , the equality will hold if and only if, the columns of $B_n X_c$ are in the column space of X_c ($B_n X_c = X_c G$, for some $p \times p$ matrix G) .

The preceeding work indicates that the use of weights $b_{ij} \not\equiv 1$ cannot increase the efficiency of the estimate over the unweighted case. It may be possible to choose weights so as to lose little or no efficiency and yet gain in some other respect. This matter needs further study.

By using weights $b_{ij} = 0$ or 1 we can reduce the number of terms in the dispersion function D by reducing the number of comparisons.

This may have computational advantages. Some examples in a later section show that there need not be a loss in efficiency. With such a choice of weights the dispersion function depends on "restricted" ranks of the residuals Z_i. Specifically, using form (2.2),

$$D = \sum_{i=1}^{n} (2R_i - (\#C_i + 1))Z_i ,$$

where $C_i = \{j: b_{ij} = 1 , 1 \leq j \leq n , j \neq i\}$, the number of elements in C_i is denoted by $\#C_i$ and R_i is the rank of Z_i in the set $\{Z_j: j \in C_i\}$.

The weights b_{ij} are associated directly with the observations (not on the ordered observations) and can be chosen to depend on the design matrix $\underset{\sim}{X}$. It may be possible to reduce the effect of high leverage points in the design matrix with suitable choice of weights.

When $\underset{\sim}{X}_c' \underset{\sim}{X}_c$ is nearly singular, the use of weights can reduce the effects of multicolinearity.

One possible approach to setting the weights is to assign a weight w_i for each observation $i = 1, \ldots, n$ and the use $b_{ij} = w_i w_j$, $i \neq j$. Assume that $\sum_{i=1}^{n} w_i = 1$ for simplicity. Define an $n \times n$ diagonal matrix $\underset{\sim}{W}$ to have i^{th} diagonal element w_i and a vector $\underset{\sim}{w} = (w_1, \ldots, w_n)'$. Then the weight matrix is

$$\begin{aligned} \underset{\sim}{B}_n &= \underset{\sim}{W} - \underset{\sim}{w}\,\underset{\sim}{w}' \\ &= (I_n - \underset{\sim}{W}\,\underset{\sim}{J}_n)\underset{\sim}{W}(\underset{\sim}{I}_n - \underset{\sim}{J}_n \underset{\sim}{W}) \\ &= (I_n - \underset{\sim}{W}\,\underset{\sim}{J}_n)(\underset{\sim}{W} - \underset{\sim}{w}\,\underset{\sim}{w}')(\underset{\sim}{I}_n - \underset{\sim}{J}_n \underset{\sim}{W}) . \end{aligned}$$

Note that $\underset{\sim}{B}_n \underset{\sim}{X}$ has $(i, j)^{th}$ element $w_i(x_{ij} - \overline{x}_j^*)$, where $\overline{x}_j^* = \sum_i w_i x_{ij}$ is a weighted average of the j^{th} column of $\underset{\sim}{X}$. $\underset{\sim}{B}_n \underset{\sim}{X}$ is then a "weighted", "centered" design matrix. Overall, this approach seems to be worth further consideration because it is a simpler task to assign n individual weights than to deal with $\binom{n}{2}$ pairwise weights.

8. EXAMPLES

Example 1 Suppose there are n_1 observations (Group 1) following the model

$$Y_i = \beta_o + \beta_1 x_{io} + \cdots + \beta_p x_{ip} + e_i^* , \quad i = 1, \ldots, n_1$$

and another n_2 observations (Group 2) following the model

$$Y_i = \beta_o^* + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + e_i^* , \quad i = n_1 + 1, \ldots, n_1 + n_2 = n .$$

Note the different intercept parameters and different error variables with possibly different distributions. Actually, the possibility of different error distributions has not been covered in the work of this paper, but the necessary modifications should be possible. Suppose the goal is to estimate $\beta_1, \ldots, \beta_p$. In some situations it may not be appropriate to compare observations in different groups. The groups may refer to different types of people, locations, times or some other blocking variable. In such a case, the between group comparisons can be excluded by using

$$b_{ij} = \begin{cases} 1/n_1 & \text{if } 1 \le i < j \le n_1 \\ 1/n_2 & \text{if } n_1 + 1 \le i < j \le n_1 + n_2 \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$D = (1/n_1) \sum_{i<j\leq n_1} |Z_i - Z_j| + (1/n_2) \sum_{n_1<i<j} |Z_i - Z_j|$$

and the weight matrix is

$$\underset{\sim}{B}_n = \begin{pmatrix} \underset{\sim}{I}_{n_1} - (1/n_1)\underset{\sim}{J}_{n_1} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{I}_{n_2} - (1/n_2)\underset{\sim}{J}_{n_2} \end{pmatrix} .$$

This is an idempotent matrix and the covariance formula (6.2) reduces to

$$(8.1) \qquad \text{cov}(\hat{\beta}_1, \ldots, \hat{\beta}_p) = (1/12(\int f^2)^2)(\underset{\sim}{X}' \underset{\sim}{B}_n \underset{\sim}{X})^{-1} .$$

The usual approach to the analysis, when the error variables all have the same distribution, is to define an indicator variable for groups

$$x_{i,p+1} = \begin{cases} 0 & \text{if } 1 \leq i \leq n_1 \\ 1 & \text{if } n_1 < i \leq n_1 + n_2 \end{cases}$$

and add this term to the model. Then use $b_{ij} \equiv 1$, the unweighted dispersion function. From formula (6.3), using an augmented design matrix X^* with the additonal column,

$$\text{cov}(\hat{\beta}_1, \ldots, \hat{\beta}_p, \hat{\beta}_{p+1}) = (1/12(\int f^2)^2)(\underset{\sim c}{X^*}' \underset{\sim c}{X^*})^{-1} .$$

When the covariance matrix of $\hat{\beta}_1, \ldots, \hat{\beta}_p$ is determined in the usual way, it is found to agree exactly with (8.1). Thus both methods yield the same asymptotic covariance matrix for the estimates. The omission of the between group comparisons does not affect the efficiency of the estimates.

<u>Example 2 (One Factor Analysis of Variance)</u> Suppose there are $p + 1$ groups of observations with sample sizes $n_1, n_2, \ldots, n_{p+1}$, $n = \sum_k n_k$. The usual model is

$$Y_i = \beta_o + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + e_i$$

where

$$x_{ij} = \begin{cases} 1 & \text{if } i = n_j + 1, \ldots, n_j + n_{j+1} \\ 0 & \text{otherwise,} \end{cases}$$

for $i = 1, \ldots, n$, $j = 1, \ldots, p$. The design matrix is

$$\underset{\sim}{X} = \begin{pmatrix} \underset{\sim}{0} & \underset{\sim}{0} & \cdot & \cdot & \cdot & \underset{\sim}{0} \\ \underset{\sim}{1} & \underset{\sim}{0} & \cdot & \cdot & \cdot & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{1} & \cdot & \cdot & \cdot & \underset{\sim}{0} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \underset{\sim}{0} & \underset{\sim}{0} & \cdot & \cdot & \cdot & \underset{\sim}{1} \end{pmatrix},$$

where $\underset{\sim}{0}$ and $\underset{\sim}{1}$ are column vectors of appropriate size. This is the usual "shift" model with location parameters β_0, $\beta_0 + \beta_1, \ldots,$ $\beta_0 + \beta_p$ for the $p + 1$ groups.

Suppose that for the dispersion function D we use weights

$$b_{ij} = \begin{cases} 1 & \text{if subscripts } i \text{ and } j \text{ are from different groups} \\ 0 & \text{if subscripts } i \text{ and } j \text{ are from the same group.} \end{cases}$$

Then the weight matrix is

$$\underset{\sim}{B}_n = \begin{pmatrix} m_1\underset{\sim}{I} & -\underset{\sim}{J} & \cdot & \cdot & \cdot & -\underset{\sim}{J} \\ -\underset{\sim}{J} & m_2\underset{\sim}{I} & \cdot & \cdot & \cdot & -\underset{\sim}{J} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ -\underset{\sim}{J} & -\underset{\sim}{J} & \cdot & \cdot & \cdot & m_{p+1}\underset{\sim}{I} \end{pmatrix},$$

where $m_i = \sum_{k \neq i} n_k$ is the number of observations outside of the i^{th} group and $\underset{\sim}{I}$ and $\underset{\sim}{J}$ are appropriately sized identity and unit matrices. With this choice of weights, only the between group comparisons are used. Formula (6.2) gives the covariance matrix of $\hat{\underset{\sim}{\beta}}$ using the above $\underset{\sim}{B}_n$ matrix.

As an alternative, all comparisons could be made with $b_{ij} \equiv 1$. Then the weight matrix is a multiple of $\underset{\sim}{B}^*_n = \underset{\sim}{I}_n - (1/n)\underset{\sim}{J}_n$. It can be shown that the covariance matrix of $\hat{\underset{\sim}{\beta}}$ is the same using $\underset{\sim}{B}^*_n$ as it is using $\underset{\sim}{B}_n$. This shows there is no loss of efficiency in dropping the within group comparisons. In fact, the estimates $\hat{\beta}$ are identical for these two cases since the terms involving within group comparisons do not involve any parameters and cannot affect the minimization process.

Another interesting point will be illustrated with a special case. Suppose $p = 3$ and there are equal sample sizes $n_k = m$, $k = 1, 2, 3, 4$. Suppose that comparisons are only made between adjacent groups in the dispersion function; that is groups 1 vs 2, 2 vs 3 and 3 vs 4. The corresponding weight matrix is

$$\underset{\sim}{B} = \begin{pmatrix} m\underset{\sim}{I} & -\underset{\sim}{J} & \underset{\sim}{0} & \underset{\sim}{0} \\ -\underset{\sim}{J} & 2m\underset{\sim}{I} & -\underset{\sim}{J} & \underset{\sim}{0} \\ \underset{\sim}{0} & -\underset{\sim}{J} & 2m\underset{\sim}{I} & -\underset{\sim}{J} \\ \underset{\sim}{0} & \underset{\sim}{0} & -\underset{\sim}{J} & m\underset{\sim}{I} \end{pmatrix} .$$

By direct computation,

$$(\underset{\sim}{X}'\underset{\sim}{B}\underset{\sim}{X})^{-1}(\underset{\sim}{X}'\underset{\sim}{B}\underset{\sim}{B}\underset{\sim}{X})(\underset{\sim}{X}'\underset{\sim}{B}\underset{\sim}{X})^{-1} = (1/m)\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} .$$

If all comparisons are made, $\underset{\sim}{B}^* = \underset{\sim}{I} - (1/4m)\underset{\sim}{J}$ and

$$(\underset{\sim}{X}'\ \underset{\sim}{B}^*\ \underset{\sim}{X})^{-1}(\underset{\sim}{X}'\ \underset{\sim}{B}^*\ \underset{\sim}{B}^*\ \underset{\sim}{X})(\underset{\sim}{X}'\ \underset{\sim}{B}^*\ \underset{\sim}{X})^{-1} = (1/m)\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}.$$

Thus the approach with adjacent group comparisons yields the same covariance matrix as the usual approach with all comparisons. This suggests that there is considerable redundancy in making all comparisons.

The possibility of restricting the number of comparisons, without a loss of efficiency, may be especially useful in more complicated fixed effect designs. Such designs can be viewed in terms of a one-way layout with the parameters of interest being contrasts in the location parameters of the groups.

9. APPENDIX

Proof of Theorem 4.1

Let $w_i = \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$, $i = 1, \ldots, n$, and $\bar{w} = \sum_i w_i/n$. Let $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_p)'$ be a fixed $p \times 1$ vector and without loss of generality in what is to follow take $\underset{\sim}{\theta}' \underset{\sim}{\theta} = 1$. Consider a linear combination $U = U(\underset{\sim}{\beta}) = \underset{\sim}{\theta}' \underset{\sim}{T}(\underset{\sim}{\beta}) = \sum_{i<j} h_{ij} \phi(Z_i, Z_j)$, where $h_{ij} = \theta_1 a_{ij}(1) + \cdots + \theta_p a_{ij}(p)$, $1 \le i < j \le n$. Let

$$h_{i.} = \sum_{j=i+1}^{n} h_{ij}, \quad h_{.j} = \sum_{i=1}^{j-1} h_{ij}, \quad h_{n.} = 0, \quad h_{.1} = 0, \quad h_{..} = \sum_{i<j} h_{ij},$$

$$H_i = h_{.i} - h_{i.} = \theta_1 A_i(1) + \cdots + \theta_p A_i(p).$$

The theorem will follow if it is shown that $U(\underset{\sim}{0})$ is asymptotically normal with mean $\sum_{i=1}^{n} H_i(w_i - \bar{w})(\int f^2) + (h_{..}/2)$ and variance $\sum_{i=1}^{n} H_i^2/12$ for any choice of $\underset{\sim}{\theta}$. To show this, Theorem 4 of Sievers (1978) can be directly applied once the four assumptions there are verified.

The first assumption requires $\max\limits_{1 \le i \le n} (w_i - \bar{w})^2 \to 0$. This follows from Assumption (A_3) .

The second assumption requires $\sum_i (w_i - \bar{w})^2 \to \sigma_w^2 > 0$. This follows from Assumption (A_4) using $\sum_i (w_i - \bar{w})^2 = \underset{\sim}{\theta}' \underset{\sim}{X}_c' \underset{\sim}{X}_c \underset{\sim}{\theta} /n$.

The third assumption requires $\sum_i H_i^2 / \max\limits_i H_i^2 \to \infty$.

To examine this, write

$$\frac{\sum_i H_i^2}{\max_i H_i^2} = \frac{\sum_i [\theta_1 A_i(1) + \cdots + \theta_p A_i(p)]^2}{\sum_i A_i^2(1) + \cdots + \sum_i A_i^2(p)} \cdot \frac{\sum_i A_i^2(1) + \cdots + \sum_i A_i^2(p)}{\max_i [\theta_1 A_i(1) + \cdots + \theta_p A_i(p)]^2}$$

$$= F_1 F_2 \text{ say .}$$

Now

$$F_1 = \frac{\underset{\sim}{\theta}' \underset{\sim}{V}_n \underset{\sim}{\theta}}{tr(\underset{\sim}{V}_n)}$$

$$\geq \min_{\underset{\sim}{\theta} \ni \underset{\sim}{\theta}'\underset{\sim}{\theta} = 1} \left[\frac{\underset{\sim}{\theta}' \underset{\sim}{V}_n \underset{\sim}{\theta}}{tr(\underset{\sim}{V}_n)} \right]$$

$$= \frac{\lambda_{1n}}{\lambda_{1n} + \cdots + \lambda_{pn}} ,$$

where $\lambda_{1n} \leq \lambda_{2n} \leq \cdots \leq \lambda_{pn}$ are the ordered eigenvalues of $\underset{\sim}{V}_n$. Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_p$ denote the ordered eigenvalues of $\underset{\sim}{V}$. Then by Assumption (A_5)

$$\frac{\lambda_{1n}}{\lambda_{1n} + \cdots + \lambda_{pn}} = \frac{\gamma_n \lambda_{1n}}{\gamma_n \lambda_{1n} + \cdots + \gamma_n \lambda_{pn}} \to \frac{\lambda_1}{\lambda_1 + \cdots + \lambda_p} > 0$$

as $n \to \infty$ and this shows F_1 is bounded from below away from zero.

Now with the Cauchy-Schwarz inequality used in the denominator

$$F_2 \geq \frac{\sum_i A_i^2(1) + \cdots + \sum_i A_i^2(p)}{\underset{\sim}{\theta}' \underset{\sim}{\theta} \max_i [A_i^2(1) + \cdots + A_i^2(p)]}$$

$$\geq \frac{\sum_i A_i^2(1) + \cdots + \sum_i A_i^2(p)}{p \max\{A_i^2(1), \ldots, A_i^2(p)\}}$$

$$\geq \frac{1}{p} \min \left\{ \frac{\sum_i A_i^2(1)}{\max_i A_i^2(1)}, \ldots, \frac{\sum_i A_i^2(p)}{\max_i A_i^2(p)} \right\} .$$

Then, with Assumption (A_1), it follows that $F_2 \to \infty$ as $n \to \infty$. The asymptotic behavior of F_1 and F_2 then guarantee that the third assumption is met.

The fourth assumption requires that $\sum_{i<j} h_{ij}^2 / \sum_i H_i^2 \to 0$.

Now

$$\frac{\sum_{i<j} h_{ij}^2}{\sum_i H_i^2} = \frac{\sum_{i<j}[\theta_1 a_{ij}(1) + \cdots + \theta_p a_{ij}(p)]^2}{\sum_i A_i^2(1) + \cdots + \sum_i A_i^2(p)} \quad \frac{\sum_i A_i^2(1) + \cdots + \sum_i A_i^2(p)}{\sum_i[\theta_1 A_i(1) + \cdots + \theta_p A_i(p)]^2}$$

$$= G_1 G_2 \text{ say .}$$

Now using the Cauchy-Schwarz inequality in the numerator

$$G_1 \leq \frac{\underset{\sim}{\theta}' \underset{\sim}{\theta} \; [\sum_{i<j} a_{ij}^2(1) + \cdots + \sum_{i<j} a_{ij}^2(p)]}{\sum_i A_i^2(1) + \cdots + \sum_i A_i^2(p)}$$

$$\leq \frac{\sum_{i<j} a_{ij}^2(1)}{\sum_i A_i^2(1)} + \cdots + \frac{\sum_{i<j} a_{ij}^2(p)}{\sum_i A_i^2(p)} .$$

Then with Assumption (A_2), it follows that $G_1 \to 0$ as $n \to \infty$.

Now $G_2 = 1/F_1$ and F_1 was shown to be bounded away from zero. Thus G_2 is bounded and with the behavior of G_1 the fourth assumption is met.

<u>Proof of Lemma 5.1</u>

It is sufficient to show that the result holds for each coordinate of $\underset{\sim}{R}(\underset{\sim}{\Delta})$. The first coordinate, say $\underset{\sim}{R_1}(\underset{\sim}{\Delta})$ will be considered. Let $t_{ij}(\underset{\sim}{\Delta}) = \Delta_1(x_{j1} - x_{i1})/\sqrt{n} + \cdots + \Delta_p(x_{jp} - x_{ip})/\sqrt{n}$. Then $R_1(\underset{\sim}{\Delta})$ can be written

$$R_1(\underset{\sim}{\Delta}) = n^{-3/2}[\textstyle\sum_{i<j} a_{ij}(1)\, W_{ij} + (\int f^2)\sum_{i<j} a_{ij}(1) t_{ij}(\underset{\sim}{\Delta})\ ] ,$$

where

$$W_{ij} = \phi(Z_i\, , Z_j) - \phi(Y_i\, , Y_j)$$

$$= \begin{cases} +1 & \text{if} \quad t_{ij}(\underset{\sim}{\Delta}) < Y_j - Y_i < 0 \\ -1 & \text{if} \quad 0 < Y_j - Y_i < t_{ij}(\underset{\sim}{\Delta}) \\ 0 & \text{otherwise} . \end{cases}$$

Actually W_{ij} can be $\pm$ 1/2 when ties occur but this will be ignored since such events have zero probability.

In both cases $t_{ij}(\underset{\sim}{\Delta}) > 0$ and $t_{ij}(\underset{\sim}{\Delta}) < 0$, $E(W_{ij}) = G(0) - G(t_{ij}(\underset{\sim}{\Delta})) = -t_{ij}(\underset{\sim}{\Delta})\, g(\xi_{ij}(\underset{\sim}{\Delta}))$, where $|\xi_{ij}(\underset{\sim}{\Delta})| \le |t_{ij}(\underset{\sim}{\Delta})|$. Then using $g(0) = \int f^2$,

$$E(R_1(\underset{\sim}{\Delta})) = n^{-3/2} \sum_{i<j} a_{ij}(1)\, t_{ij}(\underset{\sim}{\Delta})\, [g(0) - g(\xi_{ij}(\underset{\sim}{\Delta}))] .$$

From Assumptions (A_3) and (A_6), it follows that for any $\varepsilon > 0$ $\max_{i<j} |g(0) - g(\xi_{ij}(\underset{\sim}{\Delta}))| < \varepsilon$ uniformly in $\underset{\sim}{\Delta} \in \mathcal{D}$ for n sufficiently large. Further, noting that $\sum_{i<j}(x_{jk} - x_{ik})(x_{j\ell} - x_{i\ell})/\binom{n}{2} = \sum_i (x_{ik} - \bar{x}_k)(x_{i\ell} - \bar{x}_\ell)/n$, write $\sum_{i<j} t_{ij}^2(\underset{\sim}{\Delta})/\binom{n}{2} = \underset{\sim}{\Delta}'\, \underset{\sim}{X}_c'\, \underset{\sim}{X}_c\, \underset{\sim}{\Delta}/n$.

Then, with use of the Cauchy-Schwarz inequality,

$$|E(R_1(\underset{\sim}{\Delta}))| \le \varepsilon\, n^{-2} \binom{n}{2} \{ \sum_{i<j} a_{ij}^2(1)/\binom{n}{2} \}^{1/2} \quad \{\underset{\sim}{\Delta}'\, \underset{\sim}{X}_c'\, \underset{\sim}{X}_c\, \underset{\sim}{\Delta}/n\}^{1/2} .$$

With Assumptions (A_4) and (A_7) and the fact that ε is arbitrary, it must be that $E(R_1(\underset{\sim}{\Delta})) \to 0$ as $n \to \infty$, uniformly in $\underset{\sim}{\Delta} \in \mathcal{D}$.

The variance of $R_1(\underset{\sim}{\Delta})$ is

$$\mathrm{Var}(R_1(\underset{\sim}{\Delta})) = n^{-3} \sum_{i<j} a_{ij}^2(1)\, \mathrm{Var}(W_{ij})$$

$$+ n^{-3} \sum_{i<j} \sum_{\substack{k<\ell \\ (i,j) \neq (k,\ell)}} a_{ij}(1)\, a_{k\ell}(1)\, \mathrm{cov}(W_{ij}, W_{k\ell}) .$$

Using Assumptions (A_3) and (A_7) , it can be shown that $\mathrm{Var}(R_1(\underset{\sim}{\Delta})) \to 0$ as $n \to \infty$, uniformly in $\underset{\sim}{\Delta} \in \mathcal{D}$. With the mean and variance tending to zero, the lemma follows.

Proof of Theorem 5.1

It is sufficient to show that the result holds for each coordinate of $\underset{\sim}{R}(\underset{\sim}{\Delta})$. Suppose $R_1(\underset{\sim}{\Delta})$ is considered. $R_1(\underset{\sim}{\Delta})$ and $t_{ij}(\underset{\sim}{\Delta})$ are given in the preceeding proof.

Let $\varepsilon > o$, $\varepsilon' > 0$ be given.

From Assumptions (A_4) and (A_7) there exists a bound B_o such that

$$(1/2)[\sum_{i<j} a_{ij}^2(1)/\binom{n}{2}]^{1/2} \{\sum_{k=1}^{p} [\sum_{i=1}^{n} (x_{ik} - \bar{x}_k)^2/n]^{1/2}\} \leq B_o$$

Now choose δ so that $3\, g(0) B_o\, \delta < \varepsilon/3$. Then partition $\mathcal{D}$ into closed subsets $\mathcal{D}_1, \ldots, \mathcal{D}_M$ say, so that $\underset{\sim}{\Delta}, \underset{\sim}{\Delta}' \in \mathcal{D}_m$ implies $|\Delta_k - \Delta'_k| \leq \delta$ for all $k = 1, \ldots, p$ and $m = 1, \ldots, M$ and that $\mathcal{D} = \bigcup_{m=1}^{M} \mathcal{D}_m$. Let $t_{ij}^U(m) = \max_{\underset{\sim}{\Delta} \in \mathcal{D}_m} t_{ij}(\underset{\sim}{\Delta})$ and $t_{ij}^L(m) = \min_{\underset{\sim}{\Delta} \in \mathcal{D}_m} t_{ij}(\underset{\sim}{\Delta})$.

Then from Assumptions (A_4) and (A_7), it can be shown that

$$n^{-3/2} \sum_{i<j} a_{ij}(1) |t_{ij}^U(m) - t_{ij}^L(m)| \leq B_o \delta .$$

For $1 \leq i < j \leq n$ and $m = 1, \ldots, M$, define random variables

$$S_{ij}(m) = \begin{cases} 1 & \text{if} \quad t_{ij}^L(m) \leq Y_j - Y_i \leq t_{ij}^U(m) . \\ 0 & \text{otherwise} \end{cases}$$

and $Q_m = n^{-3/2} \sum_{i<j} a_{ij}(1)\, S_{ij}(m)$.

Under Assumptions (A_3), (A_4), (A_6), (A_7) and $\underset{\sim}{\beta} = \underset{\sim}{0}$ it can be shown that

$$E(Q_m) = n^{-3/2} \sum_{i<j} a_{ij}(1)[G(t_{ij}^U(m)) - G(t_{ij}^L(m))]$$

$$\leq 2g(0)B_o \delta$$

for all $m = 1, \ldots, M$ and for n sufficiently large.

Further, it can be shown that

$$\mathrm{Var}(Q_m) \to 0 \quad \text{as} \quad n \to \infty ,$$

for all $m = 1, \ldots, M$.

Now for each $m = 1, \ldots, M$, choose a point $\underset{\sim}{\Delta}_m \varepsilon \mathcal{D}_m$.

Then note that

$$\sup_{\underset{\sim}{\Delta} \varepsilon \mathcal{D}_m} n^{-3/2} T_1(\underset{\sim}{\Delta}/\sqrt{n}) - T_1(\underset{\sim}{\Delta}_m/\sqrt{n})| \leq Q_m$$

for each $m = 1, \ldots, M$. Further, by Lemma 5.1,

$$P(|R_1(\underset{\sim}{\Delta}_m)| \geq \varepsilon/3) \leq \varepsilon'/2M$$

for each $m = 1, \ldots, M$ and for n sufficiently large.

Putting some pieces together, for each $m = 1, \ldots, M$ and for n sufficiently large

$$\sup_{\underset{\sim}{\Delta}\varepsilon \mathcal{D}_m} |R_1(\underset{\sim}{\Delta}) - R_1(\underset{\sim}{\Delta}_m)| \le \sup_{\underset{\sim}{\Delta}\varepsilon \mathcal{D}_m} n^{-3/2} |T_1(\underset{\sim}{\Delta}/\sqrt{n}) - T_1(\underset{\sim}{\Delta}_m/\sqrt{n})|$$

$$+ \sup_{\underset{\sim}{\Delta}\varepsilon \mathcal{D}_m} n^{-3/2} g(0) \sum_{i<j} a_{ij}(1) |t_{ij}(\underset{\sim}{\Delta}) - t_{ij}(\underset{\sim}{\Delta}_m)|$$

$$\le Q_m + g(0) B_o \delta$$

$$= Q_m - E(Q_m) + E(Q_m) + g(0) B_o \delta$$

$$\le Q_m - E(Q_m) + \varepsilon/3 \quad .$$

Further, for each $m = 1, \ldots, M$ and for n sufficiently large

$$P(\sup_{\underset{\sim}{\Delta}\varepsilon \mathcal{D}_m} |R_1(\underset{\sim}{\Delta})| \ge \varepsilon)$$

$$\le P(\sup_{\underset{\sim}{\Delta}\varepsilon \mathcal{D}_m} |R_1(\underset{\sim}{\Delta}) - R_1(\underset{\sim}{\Delta}_m)| + R_1(\underset{\sim}{\Delta}_m) \le \varepsilon)$$

$$\le P(Q_m - E(Q_m) + \varepsilon/3 + R_1(\underset{\sim}{\Delta}_m) \ge \varepsilon)$$

$$\le P(Q_m - E(Q_m) \ge \varepsilon/3) + P(|R_1(\underset{\sim}{\Delta}_m)| \ge \varepsilon/3)$$

$$\le (9/\varepsilon^2)\, \mathrm{Var}(Q_m) + \varepsilon'/2M$$

$$\le \varepsilon'/2M + \varepsilon'/2M$$

$$= \varepsilon'/M \quad .$$

Finally, for n sufficently large

$$P(\sup_{\underset{\sim}{\Delta}\in\mathcal{D}} |R_1(\underset{\sim}{\Delta})| \geq \varepsilon) \leq \sum_{m=1}^{M} P(\sup_{\underset{\sim}{\Delta}\in\mathcal{D}_m} |R_1(\underset{\sim}{\Delta})| \geq \varepsilon)$$

$$\leq \sum_{m=1}^{M} \varepsilon'/M$$

$$= \varepsilon'$$

and the proof is completed.

Proof of Lemma 6.2

When $(1/n^2)\underset{\sim}{C}_n$ in the expression $\underset{\sim}{R}(\underset{\sim}{\Delta})$ is replaced by $\underset{\sim}{C}$, the proof of this lemma is routine.

Proof of Lemma 6.3

Note that

$$\frac{\partial Q(\underset{\sim}{\Delta})}{\partial \underset{\sim}{\Delta}} - \frac{\partial D^*(\underset{\sim}{\Delta})}{\partial \underset{\sim}{\Delta}} = 2n^{-3/2}[\underset{\sim}{T}(\underset{\sim}{\Delta}/\sqrt{n}) - \underset{\sim}{T}(\underset{\sim}{0}) - n^{3/2}\, \underset{\sim}{C}\, \underset{\sim}{\Delta}]$$

and use Lemma 6.1.

Proof of Lemma 6.4

With Lemma 6.3, the proof finishes exactly as in Jaeckel (1972), page 1454.

Proof of Theorem 6.1

Again use exactly the argument of Jaeckel (1972), page 1454, along with Lemma 6.4.

REFERENCES

Adichie, J. N. (1978), "Rank Tests of Sub-Hypotheses in the General Linear Regression," Annals of Statistics, 6, 1012-1026.

Hettmansperger, T. P. and McKean, J. W. (1977), "A Robust Alternative Based on Ranks to Least Squares in Analyzing Linear Models," Technometrics, 19, 275-284.

Hettmansperger, T. P. and McKean, J. W. (1978a), "Statistical Inference Based on Ranks," Psychometrika, 43, 69-79.

Jaeckel, L. A. (1972), "Estimating Regression Coefficients By Minimizing the Dispersion of the Residuals," Annals of Mathematical Statistics, 43, 1449-1458.

Jurečková, J. (1969), "Asymptotic Linearity of a Rank Statistic in Regression Parameter," Annals of Mathematical Statistics, 40, 1889-1900.

Jurečková, J. (1971), "Nonparametric Estimate of Regression Coefficients," Annals of Mathematical Statistics, 42, 1328-1338.

Koul, H. L. (1970), "A Class of ADF Tests For Subhypothesis in the Multiple Linear Regression," Annals of Mathematical Statistics, 41, 1273-1281.

Koul, H. L. (1971), "Asymptotic Behavior of a Class of Confidence Regions Based on Ranks in Regression," Annals of Mathematical Statistics, 42, 466-476.

McKean, J. W. and Hettmansperger, T. P. (1976), "Tests of Hypotheses Based on Ranks in the General Linear Model," Communications in Statistics - Theory and Methods - A, 8, 693-709.

McKean, J. W. and Hettmansperger, T. P. (1978b), "A Robust Analysis of the General Linear Model Based on One Step R-Estimates " Biometrika, 65, 571-579.

Sen, P. K. and Puri, M. L. (1977), "Asymptotically Distribution-Free Aligned Rank Order Tests for Composite Hypotheses for General Multivariate Linear Models," Z. Wahrscheinlichkeitstheorie Verw. Gebiete, 39, 175-186.

Sievers, G. L. (1978), "Weighted Rank Statistics for Simple Linear Regression," Journal of the American Statistical Association, 73, 628-631.

Wainer, H. and Thissen, D. (1976), "Three Steps Towards Robust Regression," Psychometrika, 41, 9-34.